Serial No. 10/006,492

AMENDMENTS TO THE CLAIMS:

A listing of the entire set of claims (including amendments to the claims) is submitted herewith. The listing of claims will replace all prior versions, and listing of the claims in the application.

1 - 40. (Canceled)

41. (Withdrawn)   A method comprising:

defining a sequential pattern of biopolymer sequence segments, the pattern comprising a similar segment and a dissimilar segment;

comparing a first biopolymer sequence to a reference to identify similar and dissimilar segments in the first sequence; and

determining if the similar and dissimilar segments of the first biopolymer sequence match the defined sequential pattern.

42. (Withdrawn)   The method of claim 41 in which the comparing and the determining are concurrent.

43. (Withdrawn)   The method of claim 41 in which the reference comprises a second biopolymer sequence.

44. (Withdrawn)   The method of claim 41 in which the reference comprises a sequence profile.

45. (Withdrawn)   The method of claim 41 further comprising repeating the comparing and determining for a plurality of sequences.

46. (Withdrawn)   The method of claim 41 further comprising repeating the comparing and determining such that multiple combinations of sequences selected from a plurality of sequences are compared.

2

Serial No. 10/006,492

47. (Withdrawn)   The method of claim 46 in which the plurality of sequences comprises sequences from different species of the same phyla.

48. (Withdrawn)   The method of claim 47 in which the plurality of sequences comprises sequences from different mammalian species.

49. (Withdrawn)   The method of claim 47 in which each of the multiple combinations includes sequences from different species.

50. (Withdrawn)   The method of claim 41 in which the determining comprises identifying a value that evaluates the matching to the defined sequential pattern.

51. (Withdrawn)   The method of claim 46 further comprising ranking the combinations based on the identified value.

52. (Withdrawn)   The method of claim 41 further comprising, if the similar and dissimilar segments of the first biopolymer sequence match the defined sequential pattern, assaying a biopolymer that comprises one of the segments of the first biopolymer sequence for an activity.

53. (Withdrawn)   The method of claim 52 in which the biopolymer comprises the similar segment.

54. (Withdrawn)   The method of claim 52 in which the biopolymer comprises the first polymer sequence.

55. (Withdrawn)   A method comprising:
    evaluating sets, each set comprising a first sequence from sequences of a first species and a second sequence from sequences of a second species, the

3

Serial No. 10/006,492

evaluating comprising

    (i) comparing the first and second sequence of each set to identify similar and dissimilar segments; and

    (ii) returning a value indicative of the match between the similar and dissimilar segments of the set and a defined pattern of similarity and dissimilarity; and

    identifying sets which return values that exceed a threshold.

56. (Withdrawn)   The method of claim 55 in which the first species is a eukaryotic species.

57. (Withdrawn)   The method of claim 56 in which the first species is a vertebrate species.

58. (Withdrawn)   The method of claim 57 in which the first species is a mammalian species.

59. (Withdrawn)   The method of claim 58 in which the first species is a human.

60. (Withdrawn)   The method of claim 58 in which the second species is a mammalian species.

61. (Withdrawn)   The method of claim 55 in which the similar segment is between processing sites.

62. (Withdrawn)   The method of claim 55 in which the similar segment is adjacent to a processing site.

4

Serial No. 10/006,492

63. (Withdrawn)   The method of claim 61 in which the dissimilar segment is outside the processing sites.

64. (Withdrawn)   The method of claim 62 in which the processing site is a protease cleavage site.

65. (Withdrawn)   A method comprising:

a) comparing a query sequence to each candidate sequence of a plurality of candidate sequences by a method comprising

i) identifying a first segment in the candidate sequence and a first segment in a query sequence;

ii) determining a first measure that is a measure of the similarity between the first segments; and

iii) determining a second measure that is a measure of the similarity between segments of the query sequence and the candidate sequence, the segments being other than the first segment; and

b) identifying a selected candidate sequence from the plurality of candidate sequences, wherein a comparison of the first and second measures of the selected candidate sequence indicate at least a threshold value.

66. (Withdrawn)   The method of claim 65 in which each first segment is adjacent to a processing site.

67. (Withdrawn)   The method of claim 66 in which the processing site is a convertase processing site.

68. (Withdrawn)   The method of claim 65 in which the first segment is between a first processing site and second site that is a second processing site, a signal sequence, or a carboxy terminus.

5

Serial No. 10/006,492

69. (Withdrawn)   The method of claim *65* in which the identifying comprises aligning the query sequence and the candidate sequence.

70. (Withdrawn)   The method of claim 69 in which the aligning comprises maximizing local alignments.

71.   (Withdrawn) An article of machine-readable media having encoded thereon software configured to cause a processor to:

a) compare a query sequence to each candidate sequence of a plurality of candidate sequences by a method comprising

i) identifying a first segment in the candidate sequence and a first segment in a query sequence;

ii) determining a first measure that is a measure of the similarity between the first segments; and

iii) determining a second measure that is a measure of the similarity between segments of the query sequence and the candidate sequence, the segments being other than the first segment; and

b) identify a selected candidate sequence from the plurality of candidate sequences, wherein a comparison of the first and second measures of the selected candidate sequence indicate at least a threshold extent of localized similarity.

72. (Currently Amended)   A method for identifying biopolymer sequences characterized by a topological pattern of match states, the method comprising the steps of:

constructing a statistical model comprising a hidden Markov Model of a set of known sequences characterized by a topological pattern of match states, each match state characterized by a scoring matrix, wherein the scoring matrix for a first match state defines a state of similarity and the scoring matrix for a second match state defines a state of dissimilarity; the model comprising one or more modules of nodes,

comparing the topological pattern of match states of the biopolymer sequences

6

Serial No. 10/006,492

to the topological pattern of match states of the set of known sequences <u>and comparing</u>
<u>the scoring matrices of match states to determine the state of similarity or the state of</u>
<u>dissimilarity of the biopolymer sequences and the set of known sequences;</u> and

identifying the biopolymer sequences <u>based on the state of similarity or the state</u>
<u>of dissimilarity with the set of known sequences.</u>

73. (Previously Presented) The method of claim 72, wherein the step of
constructing the model comprises:

determining the topological pattern of match states of the set of known
sequences;

preparing at least one module of nodes for each match state; and

linking the modules of nodes to form the model.

74. (Previously Presented) The method of claim 73, wherein the step of
preparing the modules of nodes comprises:

programming the modules of nodes against a training set of data objects
characteristic of the topology pattern of match states of the set of known sequences;
and

tuning the nodes in an iterative process until the modules encompass the training
set of data objects.

75. (Currently Amended) The method of claim 74, wherein the step of
programming the modules of nodes comprises defining <u>the</u>a scoring matrix for each
match state.

76. (Canceled)

77. (Currently Amended) The method of claim <u>75</u>74, wherein the scoring matrix
defining a state of dissimilarity is a function of the scoring matrix defining a state of
similarity.

7

Serial No. 10/006,492

78. (Currently Amended)' The method of claim 7574, wherein the scoring matrix defining a state of dissimilarity is a function of the arithmetic inverse of the scoring matrix defining a state of similarity.

79. Cancelled)

80. (Previously Presented) The method of claim 72, wherein the set of known sequences consists of two sequences.

81. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises at least three sequences.

82. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises amino acid sequences.

83. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises nucleic acid sequences.

84. (Previously Presented) The method of claim 72, wherein one or more nodes represent an insertion at a first position in the set of known sequences.

85. (Previously Presented) The method of claim 72, wherein one or more nodes represent a deletion at a second position in the set of known sequences.

86. (Previously Presented) The method of claim 72, wherein each node represents a distribution of monomers at defined positions in the set of known sequences.

87. (Previously Presented) The method of claim 86, wherein the distribution of

8

Serial No. 10/006,492

monomers at a first node is different from the distribution of monomers at a second node.

88. (Previously Presented) The method of claim 87, wherein the distribution of monomers is a function of a scoring matrix that relates the distribution of monomers at a first node and a scoring matrix that relates the distribution of monomers at a second node.

89. (Previously Presented) The method of claim 88, wherein the scoring matrix is a function of independent probabilities of a monomer occurrence.

90. (Previously Presented) The method of claim 89, wherein the distribution $P(a,b)$ of monomers a and b, a scoring matrix $S(a,b)$, and independent probabilities of monomers, $Q(a)$ and $Q(b)$ are related such that $S(a,b) = \log(P(a,b) / (Q(a)\text{-}Q(b)))$.

91. (Currently Amended) The method of claim 72, wherein the model comprises a first module which characterizes thea match state between the set of known sequences in a first region and a second module which characterizes the match state between the set of known sequences in a second region; wherein the match states of the first and second module are different.

92. (Previously Presented) The method of claim 91, wherein the model further comprises a third module that characterizes the match state between the set of known sequences in a third region.

93. (Previously Presented) The method of claim 92, wherein the third module is positioned between the first and second module with respect to the order of the set of known sequences.

94. (Previously Presented) The method of claim 93, wherein the third module indicates similarity between a third region of each set of known sequence, and a

9

Serial No. 10/006,492

sequence profile characterized by altered scoring matrices.

95. (Previously Presented) The method of claim 94, wherein the sequence profile comprises a profile of a modification site.

96. (Previously Presented) The method of claim 95, wherein the modification site is a processing site.

97. (Previously Presented) The method of claim 96, wherein the processing site indicates a preference for at least one basic residue.

98. (Previously Presented) The method of claim 96, wherein the processing site indicates a preference for at least two basic residues.

99. (Previously Presented) The method of claim 96, wherein the processing site comprises a convertase processing site.

100. (Previously Presented) The method of claim 96, wherein the processing site comprises a secretase processing site.

101. (Previously Presented) The method of claim 72, wherein the biopolymer sequences comprise sequences from different species.

102. (Previously Presented) The method of claim 101, wherein the different species comprise mammalian species.

103. (Previously Presented) The method of claim 72, wherein the set of known sequences comprise genomic nucleic acid sequences.

104. (Previously Presented) The method of claim 72, wherein the set of known

10

Serial No. 10/006,492

sequences comprises non-coding regions.

105. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises regulatory regions.

106. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises transcriptional regulatory regions.

107. (Currently Amended)A method for identifying biopolymer sequences characterized by a topological pattern of match states, the method comprising the steps of:

constructing a statistical model comprising a hidden Markov Model of a set of known sequences characterized by a topological pattern of match states, each match state characterized by a scoring matrix, wherein the scoring matrix for a first match state defines a state of similarity and the scoring matrix for a second match state defines a state of dissimilarity, the model comprising one or more modules of nodes, each module of nodes representing a different match state, wherein the step of constructing the model comprises programming the modules of nodes against a training set of data objects characteristic of the match state of the set of known sequences, tuning the nodes in an iterative process until the modules encompass the training set of data objects, and linking the modules to form the model;

comparing the topological pattern of match states of the biopolymer sequences to the topological pattern of match states of the set of known sequences and comparing the scoring matrices for each match state to determine the state of similarity or the state of dissimilarity of the biopolymer sequences and the set of known sequences; and

identifying the biopolymer sequences based on the state of similarity or the state of dissimilarity with the set of known sequences.

108. (Previously Presented) The method of claim 107, wherein the step of programming the modules of nodes further comprises the step of defining a scoring

Serial No. 10/006,492

matrix for each match state.

109. (Canceled)

110. (Canceled)

111. (Currently Amended)The method of claim 108~~109~~, wherein the scoring matrix defining a state of dissimilarity is a function of the scoring matrix defining a state of similarity.

112. (Currently Amended) The method of claim 108~~109~~, wherein the scoring matrix defining the state of dissimilarity is a function of the arithmetic inverse of the scoring matrix defining a state of similarity.

113. (Canceled)

114. (Previously Presented) The method of claim 107, wherein the set of known sequences consists of two sequences.

115. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises at least three sequences.

116. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises amino acid sequences.

117. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises nucleic acid sequences.

118. (Previously Presented) The method of claim 107, wherein one or more nodes represent an insertion at a first position in the set of known sequences.

12

Serial No. 10/006,492

119. (Previously Presented) The method of claim 107, wherein one or more nodes represent a deletion at defined positions in the set of known sequences.

120. (Previously Presented) The method of claim 107, wherein each node represents a distribution of monomers at defined positions in the set of known sequences.

121. (Previously Presented) The method of claim 120, wherein the distribution of monomers at a first node is different from the distribution of monomers at a second node.

122. (Previously Presented) The method of claim 121, wherein the distribution of monomers is a function of a scoring matrix that relates the distribution of monomers at a first node and a scoring matrix that relates the distribution of monomers at a second node.

123. (Previously Presented) The method of claim 122, wherein the scoring matrix is a function of independent probabilities of a monomer occurrence.

124. (Previously Presented) The method of claim 123, wherein the distribution $P(a,b)$ of monomers a and b, a scoring matrix $S(a,b)$, and independent probabilities of monomers, $Q(a)$ and $Q(b)$ are related such that $S(a,b) = \log(P(a,b) / (Q(a) \sim Q(b)))$.

125. (Previously Presented) The method of claim 107, wherein the model comprises a first module which characterizes a match state between the set of known sequences in a first region and a second module which characterizes a match state between the set of known sequences in a second region; wherein the match states of the first and second modules are different.

13

Serial No. 10/006,492

126. (Previously Presented) The method of claim 125, wherein the model further comprises a third module that characterizes a match state between the set of known sequences in a third region.

127. (Previously Presented) The method of claim 126, wherein the third module is positioned between the first and second module with respect to the order of the set of known sequences.

128. (Previously Presented) The method of claim 127, wherein the third module indicates similarity between a third region of each set of known sequence, and a sequence profile characterized by altered scoring matrices.

129. (Previously Presented) The method of claim 128, wherein the sequence profile comprises a profile of a modification site.

130. (Previously Presented) The method of claim 129, wherein the modification site is a processing site.

131. (Previously Presented) The method of claim 130, wherein the processing site indicates a preference for at least one basic residue.

132. (Previously Presented) The method of claim 130, wherein the processing site indicates a preference for at least two basic residues.

133. (Previously Presented) The method of claim 130, wherein the processing site comprises a convertase processing site.

134. (Previously Presented) The method of claim 130, wherein the processing site comprises a secretase processing site.

14

Serial No. 10/006,492

135. (Previously Presented) The method of claim 107, wherein the biopolymer sequences comprise sequences from different species.

136. (Previously Presented) The method of claim 134, wherein the different species comprise mammalian species.

137. (Previously Presented) The method of claim 107, wherein the set of known sequences comprise genomic nucleic acid sequences.

138. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises non-coding regions.

139. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises regulatory regions.

140. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises transcriptional regulatory regions.

141. (Canceled)

142. (New) A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps of a method for identifying biopolymer sequences characterized by a topological pattern of match states, comprising of:

constructing a statistical model comprising a hidden Markov Model of a set of known sequences characterized by a topological pattern of match states, each match state characterized by a scoring matrix, wherein the scoring matrix for a first match state defines a state of similarity and the scoring matrix for a second match state defines a state of dissimilarity; the model comprising one or more modules of nodes,

15

Serial No. 10/006,492

comparing the topological pattern of match states of the biopolymer sequences to the topological pattern of match states of the set of known sequences and comparing the scoring matrices of match states to determine the state of similarity or the state of dissimilarity of the biopolymer sequences and the set of known sequences; and

identifying the biopolymer sequences based on the state of similarity or the state of dissimilarity with the set of known sequences.

16